

多変量解析 <初歩の初歩>に入門(2)

やっと多変量解析?

なかもとけいこ
kenakamoto@nifty.com

多変量解析とは?

□ その名のとおり、たくさんの変量をまとめて分析するための統計的手法の総称。ものすごく色々ある。

	(主に)量的データ	(主に)質的データ
外部基準あり	(重)回帰分析 正準判別分析 ロジスティック回帰分析	数量化 類 数量化 類
外部基準なし	主成分分析 因子分析 クラスター分析 多次元尺度法	数量化 類(対応分析)

これらのメタモデルとして構造方程式モデリング(SEM: 共分散構造分析)

多変量解析で得ること

- 世の中のことは大抵簡単にはできていない。たくさんの要因が組み合わさって結果が現れる。
 - 一つ一つの要因を順番に見ていっていたのではハッキリした結果は得られない。
 - しかし、ヒトはそんなに賢くないので、道具なしにたくさんのことを一度に考えることはできない。
- 多変量解析を使うと、すっきり整理して把握することができる!
- 勉強する手間の方が、データを前に悩む手間よりもたぶん少ないです。

得ることの例(1)

- 複数の説明変数を使った目的変数の予測。
 - 重回帰分析
 - 決定係数(説明率): それらの説明変数でどれくらい目的変数を予測できるか?
 - 偏回帰係数: 他の変数の影響を除去したときに、一つの変数がどれくらい目的変数と関連しているか?
 - 比喩の適切さと述部の慣習性による隠喩形式選好の予測

	Mean (SD)	Correlation (N=42)		Multiple regression analysis Standardized weights
		Met. pref.	Aptness	
Metaphor form preference	3.57 (.99)			
Aptness of the comparison	3.27 (.67)	.508**		.439**
Base conventionality	4.88 (.87)	.394**	.239	.289*
				R ² .337

R = .58

Nakamoto & Kusumi (2004: in prep.)

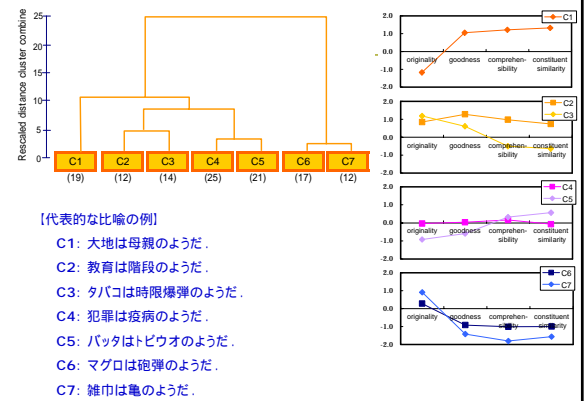
得ることの例(2)

- ケースの類似性を利用した分類(クラスタリング)
 - (階層的)クラスタ分析

■ 比喩の分類

比喩	出典	説明可能性 R ² (%)	説明類似性 n=60	説明性 R ² (%)	明白さ R ² (%)	判別率 分析結果
11 雲床-壁	yh	6.87 (2.42)	5.95 (2.06)	4.33 (2.28)	5.23 (1.92)	
90 寒波-鉄壁	kn	4.95 (2.59)	3.38 (2.10)	4.70 (1.79)	4.13 (2.04)	
62 微風-吐息	kn	6.33 (2.57)	4.78 (2.42)	4.45 (2.16)	4.51 (2.07)	
78 覆息-扉前	kn	4.95 (2.97)	3.23 (1.98)	4.52 (2.05)	4.16 (2.13)	
89 憂-季節	tk	6.78 (2.43)	4.60 (2.35)	4.98 (2.38)	5.18 (2.44)	
22 鐘論-戦争	kn	6.62 (2.06)	5.37 (2.22)	4.23 (1.87)	4.44 (2.16)	(1)
80 煙幕-閃電標弾	ao	4.16 (2.69)	3.25 (2.18)	4.70 (2.30)	4.52 (2.51)	
50 動悸-早鐘	kn	5.57 (2.80)	4.63 (2.69)	4.05 (1.74)	4.25 (2.05)	
5 激怒-噴火	kn	8.00 (1.51)	7.23 (1.94)	4.02 (2.77)	4.69 (2.41)	
21 治療-修理	kn	7.13 (2.04)	6.83 (1.67)	3.64 (2.17)	3.97 (2.24)	(1)
34 孤独-砂漠	kn	6.75 (2.40)	6.00 (1.90)	5.09 (2.24)	4.84 (2.30)	
71 氣配-定音	kn	5.63 (2.38)	5.97 (1.94)	4.30 (2.35)	3.87 (2.22)	
44 革命-地震	kn	6.03 (2.18)	4.65 (2.17)	4.55 (2.12)	4.31 (1.99)	
58 デモ-雷鳴	an	5.10 (2.45)	3.13 (2.28)	4.64 (1.91)	3.85 (1.99)	
29 記憶-音階	kn	6.56 (2.42)	5.35 (2.45)	5.06 (2.22)	4.34 (2.32)	
45 教育-階段	ao	6.79 (2.21)	4.83 (2.35)	4.77 (2.10)	4.90 (2.05)	
15 寂し-木枯らし	an	7.57 (1.99)	6.72 (2.00)	4.34 (2.64)	5.03 (2.37)	
70 マグロ-砲弾	an	3.90 (2.76)	2.55 (2.08)	4.67 (2.42)	3.97 (2.06)	

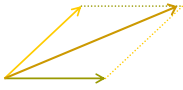
中本・橋見(2004)



相関(共分散)のベクトル表現(2)

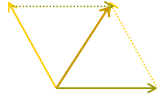
□ 合計点を表すベクトル

$r=0.7$



合計することで、値のばらつきは大きくなる。

$r=-0.5$



合計することで、値のばらつきが小さくなってしまう。

相関(共分散)のベクトル表現(3)



3次元ベクトルが3つ以上ある場合
→ 普通は3次元を使わないと表現しきれない。

が、

うまい面を発見してやると、2次元上に全部のベクトルがキレイに乗っかる場合(ベクトルが一次独立でない場合)がある。

この面を探すために解くのが相関行列の固有値, 固有ベクトル

固有値と固有ベクトル(1)

□ 多変量解析の本を読んだと、とにかくしょっちゅう出てくるコトバ。

- 主成分分析 (因子分析)
- 判別分析
- 数量化 , 類 など...

□ 固有値問題

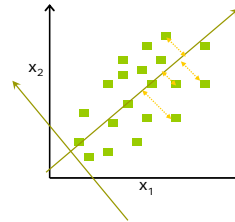
- 正方行列 A ($n \times n$) が与えられたとき、

$$Ax = \lambda x$$

- を満たす λ を固有値, x を固有ベクトルという。
- n 次の正方行列の固有値は n 個ある。
- 第1固有値が一番大きく、だんだん小さくなっていく (cf., → スクリーンプロット)

固有値と固有ベクトル(2)

□ うまい合成変量を見つけるために使う=主成分分析



データのちらばりを最も効率よく表現できる軸を探します。
→ その軸は、 x_1, x_2, \dots, x_n の合成変量
→ $f_1 = a_1x_1 + a_2x_2 + \dots + a_nx_n$ として表せる。

固有値 = 新しい軸で表せる分散の大きさに対応。

固有ベクトル = 合成得点を求めるための重みに相当

多変量解析の限界 (1)

□ 大量の数値データを縮約し、(場合によっては誤差成分をコントロールし)可視化するための技術。

- 得られる結果は、データの縮約的記述であり、それ自体が何か説明を与えるわけではない。
- 解釈は実質科学的見地 (心理学や言語学の専門家としての知識とカン) からなされるべき。

□ 結果をどこまで一般化するかは研究上の観点による。客観的基準で決められるわけではない。

- 解析結果を外挿する場合は特に慎重な判断が必要。

多変量解析の限界 (2)

□ 一般的に使用される多変量解析は複合系ではあるが、複雑系ではない。

- 基本的には線形モデルにのっとっている。
- 確率的な変動はモデルに組み込まれているが、その挙動は足し算かけ算に限られる。
- 非線形(複雑系)の現象を線形で近似的に記述していることを自覚すること。
 - 現象の全体ではなく一部を扱っているがゆえに、線形で近似できるケース
 - 非線形な現象に対し、研究者/被験者/解析者のいずれかが線形な解釈を行った結果がデータとして得られているケース

入門(2)のまとめ

- 今回は下記のことを学びました。
 - 多変量解析には色々あります。
 - 複数の変数ベクトルがくっついた行列を扱います。
 - 直接扱うのは相関行列であることも多いです。
 - 多変量正規分布という分布があります。
 - 変数はベクトルで、相関係数はベクトル間の角度(の余弦)で表せま
 - 変数に重み付けして足してやれば、新しい合成変数が得られます。これに
 - 多変量解析にも色々な点で限界はあります。
- 簡単な入門編はここで終わりです。
 - 後は自助努力あるのみ?

おまけ: 代表的な統計ソフトウェア

- SPSS (Amos, Clementine)
 - たぶん、日本の心理学界では最もユーザが多い。
 - ツールバーからコマンド選択で大抵のことは出来るのでらくちん。
 - でも、少し凝ったことをしようとしたり、一般的な使用頻度の低い分析(e.g., LOGLINEAR)をしようとすると、スクリプトを書く必要がある。
- SAS
 - スクリプトを書いて実行する方式
 - 非常に多くの分析手法をサポートしている。参考書も、和文、英文とも色々出ている。
 - スクリプトをsasウィンドウで書いていると、妙に重い...
- Statistica
 - コマンド選択式で解析可能。らくちん。
 - グラフの出力が非常にウツクシイ。
 - ユーザーズガイドが充実している。
 - けれども3.0Jは焦って日本語版をリリースしたらしくあまりデキはよくない。
- R
 - ナカモトが乗り換えを考えているソフト。ただし、私はまだ使ったことはありません。
 - なんといっても、無料なところがすごい!!!
 - Open Developmentなので、新しい解析手法が実装されるのが早い。たとえば、とっくの昔にSOM(自己組織化マップ)が実装されているし、多重比較も色々な手法を取りそろえている。
 - 最近、本もたくさんでているので、興味のある人は一緒に勉強しましょう。